TITLE OF THE INVENTION

Audio Enhancement in Coded Domain

FIELD OF THE INVENTION

[0001] The present invention relates to voice enhancement, and in particular to a method and an apparatus for enhancing a coded audio signal.

BACKGROUND OF THE INVENTION

[0002] Improved voice quality created by voice processing DSP (Digital Signal Processing) algorithms has been used to differentiate network providers. The transfer to packet networks or networks with extended tandem free operation (TFO) or transcoder free operation (TrFO) will diminish this ability to differentiate networks with traditional voice processing algorithms. Therefore, operators which have generally been responsible for maintaining speech quality for their customers are asking for voice processing algorithms to be carried out also for coded speech.

[0003] TFO is a voice standard to be deployed in the GSM (Global System for Mobile communications) and GSM-evolved 3G (Third Generation) networks. It is intended to avoid the traditional double speech encoding/decoding in mobile-to-mobile call configurations. The key inconvenience of a tandem configuration is the speech quality degradation introduced by the double transcoding. According to the ETSI listening tests, this degradation is usually more noticeable when the speech codecs are operating at low rates. Also, higher background noise level increases the degradation.

[0004] When the originating and terminating connections are using the same speech codec it is possible to transmit transparently the speech frames received from the originating MS (Mobile Station) to the terminating MS without activating the transcoding functions in the originating and terminating networks.

[0005] The key advantages of Tandem Free Operation are improvement in speech quality by avoiding the double transcoding in the network, possible savings on the inter-PLMN (Public Land Mobile Network) transmission links, which are carrying compressed speech compatible with a 16 kbit/s or 8 kbit/s sub-multiplexing scheme, including packet switched transmission, possible savings in processing power in the network equipment since the transcoding functions in the Transcoder Units are bypassed, and possible reduction in the end-to-end transmission delay.

[0006] In TFO call configuration a transcoder device is physically present in the signal path, but the transcoding functions are bypassed. The transcoding device may perform control and protocol conversion functions. In Transcoder Free Operation (TrFO), on the other hand, no transcoder device is physically present and hence no control or conversion or other functions associated with it are activated.

[0007] The level of speech is an important factor affecting the perceived quality of speech. Typically in the network side there are used automatic level control algorithms, which adjust the speech level to a certain desired target level by increasing the level of faint speech and somewhat decreasing the level of very loud voices.

[0008] These methods cannot be utilized as such in future packet networks where the speech travels in the coded format end-to-end from the transmitting device to the receiving device.

[0009] Currently the coded speech is decoded in the network and speech enhancement is carried out with linear PCM samples using traditional speech enhancement methods. After that the speech is encoded again, and transmitted to the receiving party.

[0010]

[0011] However, for example, for AMR speech codec the level control is more difficult in the lower modes due to the fact that the fixed codebook gain is no longer scalar quantized but it is vector-quantized together with the adaptive codebook gain.

SUMMARY OF THE INVENTION

[0012] It is an object of the invention to provide a method and an apparatus for enhancing a coded audio signal by means of which the above-described problems are overcome and enhancement of a coded audio signal is improved.

[0013] According to a first aspect of the invention, this object is achieved by an apparatus and a method of enhancing a coded audio signal comprising indices which represent audio signal parameters which comprise at least a first parameter representing a first characteristic of the audio signal and a second parameter, comprising:

determining a current first parameter value from an index corresponding to a first parameter;

adjusting the current first parameter value in order to achieve an enhanced first characteristic, thereby obtaining an enhanced first parameter value;

determining a current second parameter value from the index furthercorresponding to a second parameter; and

determining a new index value from a table relating index values to first parameter values and relating the index values to second parameter values, such that a new first parameter value corresponding to the new index value and a new second parameter value corresponding to the new index value substantially match the enhanced first parameter value and the current second parameter value.

[0014] According to a second aspect of the invention, this object is achieved by an apparatus and a method of enhancing a coded audio signal comprising indices which represent audio signal parameters which comprise at least a first parameter representing a first characteristic of the audio signal and a background noise parameter, comprising:

determining a current first parameter value from an index corresponding to at least a first parameter;

adjusting the current first parameter value in order to achieve an enhanced first characteristic, thereby obtaining an enhanced first parameter value;

determining a new index value from a table relating index values to at least first parameter values, such that a new first parameter value corresponding to the new index value substantially matches the enhanced first parameter value;

detecting a current background noise parameter index value; and

determining a new background noise parameter index value corresponding to the enhanced first characteristic. According to a third aspect of the invention, this object is achieved by an apparatus and a method of enhancing a coded audio signal comprising indices which represent audio signal parameters, comprising:

detecting a characteristic of an audio signal;

detecting a current background noise parameter index value; and

determining a new background noise parameter index value corresponding to the detected characteristic of the audio signal.
The invention may also be embodied as computer program product comprising portions for performing steps when the product is run on a computer.

According to an embodiment of the invention, a coded audio signal comprising speech and/or noise in a coded domain is enhanced by manipulating coded speech and/or noise parameters of an AMR (Adaptive Multi-Rate) speech codec. As a result, adaptive level control, echo control and noise suppression can be achieved in the network even if speech is not transformed into linear PCM samples, as is the case in TFO, TrFO and future packet networks.

[0015] More precisely, according to an embodiment of the invention a method for controlling the level of the AMR coded speech for all the AMR codec modes 12.2 kbit/s, 10.2 kbit/s, 7.95kbit/s, 7.40 kbit/s , 6.70 kbit/s, 5.90 kbit/s, 5.15 kbit/s and 4.75 kbit/s is described. The level of the coded speech is adjusted by changing one of the coded speech parameters, namely the quantization index of the fixed codebook gain factor in the modes 12.2 kbit/s and 7.95 kbit/s. In the rest of the modes the fixed codebook gain is jointly vector-quantized with the adaptive codebook gain, and therefore adjusting the level of the coded speech requires changing both the fixed codebook gain factor and the adaptive codebook gain (joint index).

[0016] According to the invention, a new gain index is found such that the error between the desired gain and the realized effective gain becomes minimized. The proposed level control does not cause audible artifacts.

[0017] Therefore, according to the invention, level control is enabled also in lower AMR bit rates (not only 12.2 kbit/s and 7.95 kbit/s). The level control in the AMR mode 12.2 kbit/s can be improved by taking into account the required corresponding level control for the comfort noise level.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] Fig. 1 shows a simplified model of speech synthesis in AMR.

[0019] Fig. 2 demonstrates the effect of a DTX operation on a gain manipulation algorithm with noisy child speech samples.

[0020] Fig. 3 shows a diagram illustrating a response of an adaptive codebook to a step-function.

[0021] Fig. 4 shows a non-linear 32-level quantization table of a fixed codebook gain factor in modes 12.2 kbit/s and 7.95 kbit/s.

[0022] Fig. 5 shows a diagram illustrating the difference between adjacent quantization levels in the quantization table of Fig. 4.

[0023] Fig. 6 shows a vector quantization table for an adaptive codebook gain and a fixed codebook gain in modes 10.2, 7.4 and 6.7 kbit/s.

[0024] Fig. 7 shows a vector quantization table for an adaptive codebook gain and a fixed codebook gain factor in modes 5.90 and 5.15 kbit/s.

[0025] Fig. 8 shows a diagram illustrating a change in the fixed codebook gain when the fixed codebook gain factor is changed one quantization step.

[0026] Figs. 9 and 10 show diagrams illustrating re-quantized levels of the fixed codebook gain factor.

[0027] Fig. 11 illustrates values of terms $\frac{\|\mathbf{y}\|}{\|\mathbf{z}\|}$ and $\frac{\|\mathbf{y}\|}{g_c'\|\mathbf{z}\|}$ with male speech samples.

[0028] Fig. 12 illustrates values of terms $\frac{\|\mathbf{y}\|}{\|\mathbf{z}\|}$ and $\frac{\|\mathbf{y}\|}{g_c'\|\mathbf{z}\|}$ with child speech samples.

[0029] Fig. 13 shows a flow chart illustrating a method of enhancing a coded audio signal according to the invention.

[0030] Fig. 14 shows a schematic block diagram illustrating an apparatus for enhancing a coded audio signal according to the present invention.

[0031] Fig. 15 shows a block diagram illustrating the use of fixed gain.

[0032] Fig. 16 shows a diagram illustrating a high level implementation of the invention in a media gateway.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0033] In the following, an embodiment of the present invention will be described in connection with an AMR coded audio signal comprising speech and/or noise. However, the invention is not limited to AMR coding and can be applied to any audio signal coding technique employing indices corresponding to audio signal parameters. For example, such audio signal parameters may control a level of synthesized speech. In other words, the invention can be applied to a audio signal coding technique in which an index indicating a value of an audio signal parameter controlling a first characteristic of the audio signal is transmitted as coded audio signal, in which this index may also indicate a value of an audio signal parameter controlling another audio signal characteristic such as a pitch of the synthesized speech.

[0034] The adaptive multi-rate speech codec (AMR) is presented to the extent necessary for illustrating the preferred embodiments. References 3GPP TS 26.090 V4.0.0 (2001-03), “3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory Speech Codec speech processing functions; AMR speech codec; Transcoding functions (Release 4)”, and Kondoz A. M. University of Surrey, UK, "Digital speech coding for low bit rate communications systems," chapter 6: 'Analysis-by-synthesis coding of speech,' pages 174-214. John Wiley & Sons, Chichester,1994 contain further information.The adaptive multi-rate (AMR) speech codec is based on the code-excited linear predictive (CELP) coding model. It consists of eight source codecs, or modes of operation, with bit-rates of 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 and 4.75 kbit/s. The basic encoding and decoding principles of the AMR codec are explained briefly below. In addition, the matters relevant to the parameter domain gain control are discussed in more detail.

[0035] The AMR encoding process comprises three main steps:

[0036] LPC (Linear predictive coding) analysis:

The short-term correlations between speech samples (formants) are modeled and removed by a 10th order filter. In AMR codec the LP coefficients are calculated using the autocorrelation method. The LP coefficients are further transformed to Line Spectral Pairs (LSPs) for quantization and interpolation purposes utilizing the property of LSPs having a strong correlation between adjacent subframes.

[0037] Pitch analysis (long-term prediction):

The long-term correlations between speech samples (voice periodicity) are modeled and removed by a pitch filter. The pitch lag is estimated from the perceptually weighted input speech signal by first using the computationally less expensive open-loop method. A more accurate pitch lag and pitch gain g_p is then estimated by a closed-loop analysis around the open-loop pitch lag estimate, allowing also fractional pitch lags. The pitch synthesis filter in AMR is implemented as shown in Fig. 1 using an adaptive codebook approach. That is, the adaptive codebook vector *v(n)* is computed by interpolating the past excitation signal *u(n)* at the given integer delay *k* and phase (fraction) *t*:

$$v(n)=\sum_{i=0}^{9}u(n-k-i)b_{60}(t+i\cdot 6)+\sum_{i=0}^{9}u(n-k+1+i)b_{60}(6-t+i\cdot 6), \qquad (1.1)$$
$$n=0,....,39,\quad t=0,...5,\quad k=[18,143]$$

where b_{60} is an interpolation filter based on a Hamming windowed sin(*x*)/*x* function.

[0038] Optimum excitation determination (innovative excitation search):

As shown in Fig. 1, the speech is synthesized in the decoder by adding appropriately scaled adaptive and fixed codebook vectors together and feeding it through the short-term synthesis filter. Once the parameters of the LP synthesis filter and pitch synthesis filter are found, the optimum excitation sequence in a codebook is chosen at the encoder side using an analysis-by-synthesis search procedure in which the error between the original and the synthesized speech is minimized according to a perceptually weighted distortion measure. The innovative excitation sequences consist of 10 to 2 (depending on the mode) nonzero pulses of amplitude ±1. The search procedure determines the locations of these pulses in the 40-sample subframe, as well as the appropriate fixed codebook gain g_c.

[0039] The CELP model parameters LP filter coefficients, pitch parameters, i.e. the delay and the gain of the pitch filter, and fixed codebook vector and fixed codebook gainare encoded for transmission to LSP indices, adaptive codebook index (pitch index) and adaptive codebook (pitch) gain index, and fixed codebook indices and fixed codebook gain factor index, respectively.

[0040] Next, quantization of the fixed codebook gain is explained.

[0041] To make it efficient, the fixed codebook gain quantization is performed using moving-average (MA) prediction with fixed coefficients. The MA prediction is performed on the innovation energy as follows. Let $E(n)$ be the mean-removed innovation energy (in dB) at subframe n, and given by:

$$E(n) = 10\log\left(\frac{1}{N} g_c^2 \sum_{i=0}^{N-1} c^2(i)\right) - \overline{E}, \tag{1.2}$$

where $N = 40$ is the subframe size, $c(i)$ is the fixed codebook excitation, and $\overline{E}$ (in dB) is the mean of the innovation energy (a mode-dependent constant). The predicted energy is given by:

$$\widetilde{E}(n) = \sum_{i=1}^{4} b_i \hat{R}(n-i), \tag{1.3}$$

where $[b_1\ b_2\ b_3\ b_4] = [0.68\ 0.58\ 0.34\ 0.19]$ are the MA prediction coefficients, and $\hat{R}(k)$ is the quantified prediction error at subframe k:

$$\hat{R}(k) = E(k) - \widetilde{E}(k). \tag{1.4}$$

[0042] Now, a predicted fixed codebook gain is computed using the predicted energy as in Eq. (1.2) (by substituting $E(n)$ by $\widetilde{E}(n)$ and g_c by g_c'). First, the mean innovation energy is found by:

$$E_I = 10\log\left(\frac{1}{N} \sum_{j=0}^{N-1} c^2(j)\right) \tag{1.5}$$

and then the predicted gain g_c' is found by:

$$g_c' = 10^{0.05(\widetilde{E}(n)+\bar{E}-E_I)}. \qquad (1.6)$$

[0043] A correction factor between the gain g_c and the estimated one g_c' is given by:

$$\gamma_{gc} = g_c / g_c'. \qquad (1.7)$$

[0044] The prediction error and the correction factor are related as:

$$R(n) = E(n) - \widetilde{E}(n) = 20\log(\gamma_{gc}). \qquad (1.8)$$

[0045] At the decoder, the transmitted speech parameters are decoded and speech is synthesized.

[0046] Decoding of the fixed codebook gain

In case of scalar quantization (in modes 12.2 kbit/s and 7.95 kbit/s), the decoder receives an index to a quantization table that gives the quantified fixed codebook gain correction factor $\hat{\gamma}_{gc}$.

[0047] In case of vector quantization (in all the other modes) the index gives both the quantified adaptive codebook gain $\hat{g}_p$ and the fixed codebook gain correction factor $\hat{\gamma}_{gc}$.

[0048] The fixed codebook gain correction factor gives the fixed codebook gain the same way as described above. First, the predicted energy is found by:

$$\widetilde{E}(n)=\sum_{i=1}^{4}b_i\hat{R}(n-i) \qquad (1.9)$$

and then the mean innovation energy is found by:

$$E_I=10\log\left(\frac{1}{N}\sum_{j=0}^{N-1}c^2(j)\right). \qquad (1.10)$$

[0049] The predicted gain is found by:

$$g'_c=10^{0.05\left(\widetilde{E}(n)+\overline{E}-E_I\right)}. \qquad (1.11)$$

[0050] And finally, the quantified fixed codebook gain is achieved by:

$$\hat{g}_c=\hat{\gamma}_{gc}g'_c. \qquad (1.12)$$

[0051] There are some differences between the AMR modes that are relevant to the parameter domain gain control, as listed below.

[0052] In the 12.2 kbit/s mode, the fixed codebook gain correction factor γ_{gc} is scalar quantized with 5 bits (32 quantization levels). The correction factor γ_{gc} is computed using a mean energy value $\overline{E}$=36 dB.

[0053] In the 10.2 kbit/s mode, the fixed codebook gain correction factor γ_{gc} and the adaptive codebook gain g_p are jointly vector quantized with 7 bits. The correction factor γ_{gc} is computed using a mean energy value $\overline{E}$=33 dB. Moreover, this mode includes smoothing of the fixed codebook gain. The fixed codebook gain used for synthesis in the decoder is replaced by a smoothed value of the fixed codebook gains of the previous 5 subframes. The smoothing is based on a measure of the stationarity of the short-term spectrum in the LSP (Line Spectral Pair) domain. The smoothing is performed to avoid unnatural fluctuations in the energy contour.

[0054] In the 7.95 kbit/s mode, the fixed codebook gain correction factor γ_{gc} is scalar quantized with 5 bits, as in the mode 12.2 kbit/s. The correction factor γ_{gc} is computed using a mean energy value $\overline{E}$=36 dB. This mode includes anti-sparseness processing. An adaptive anti-sparseness post-processing procedure is applied to the fixed codebook vector $c(n)$ in order to reduce perceptual artifacts arising from the sparseness of the algebraic fixed codebook vectors with only a few non-zero samples per an impulse response. The anti-sparseness processing consists of circular convolution of the fixed codebook vector with one of three pre-stored impulse responses. The selection of the impulse response is performed adaptively from the adaptive and fixed codebook gains.

[0055] In the 7.40 kbit/s mode, the fixed codebook gain correction factor γ_{gc} and the adaptive codebook gain g_p are jointly vector quantized with 7 bits, as in the mode 10.2 kbit/s. The correction factor γ_{gc} is computed using a mean energy value $\overline{E}$=30 dB.

[0056] In the 6.70 kbit/s mode, the fixed codebook gain correction factor γ_{gc} and the adaptive codebook gain g_p are jointly vector quantized with 7 bits, as in the mode 10.2 kbit/s. The correction factor γ_{gc} is computed using a mean energy value $\overline{E}$=28.75 dB. This mode includes smoothing of the fixed codebook gain, and anti-sparseness processing.

[0057] In the 5.90 and 5.15 kbit/s modes, the fixed codebook gain correction factor γ_{gc} and the adaptive codebook gain g_p are jointly vector quantized with 6 bits. The correction factor $\gamma_{gc.}$ is computed using a mean energy value $\overline{E}$=33 dB. The modes include smoothing of the fixed codebook gain and anti-sparseness processing.

[0058] In the 4.75 kbit/s mode, the fixed codebook gain correction factor γ_{gc} and the adaptive codebook gain g_p are jointly vector quantized only every 10 ms by a unique method as described in 3GPP TS 26.090 V4.0.0 (2001-03), "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory Speech Codec speech processing functions; AMR speech codec; Transcoding functions (Release 4)". This mode includes smoothing of the fixed codebook gain and anti-sparseness processing.

Discontinuous transmission (DTX)

[0059] During discontinuous transmission (DTX) only the average background noise information is transmitted at regular intervals to the decoder when speech is not present as described in 3GPP TS 26.092 V4.0.0 (2001-03), "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory Speech Codec speech processing functions; AMR speech codec; Comfort noise aspects (Release 4)". At the far-end the decoder reconstructs the background noise according to the transmitted noise parameters avoiding thus extremely annoying discontinuities in the background noise in the synthesized speech.

[0060] The comfort noise parameters, information on the level and the spectrum of the background noise are encoded into a special frame called a Silence Descriptor (SID) frame for transmission to the receive side.

[0061] For parameter domain gain control purposes, the information on the level of the background noise is of interest. If the gain level were adjusted only during speech frames, the background noise level would change abruptly at the beginning and end of noise only bursts, as illustrated in Fig. 2. The level changes in the background noise are subjectively very annoying see e.g. Kondoz A. M., University of Surrey, UK, "Digital speech coding for low bit rate communications systems," page 336, John Wiley & Sons, Chichester, 1994. The more annoying the greater the amplification or attenuation is. If the level of speech is adjusted, also the level of the background noise has to be adjusted accordingly to prevent any fluctuations in the background noise level.

[0062] At the transmitting side, the frame energy is computed for each frame marked with (Voice Activity Detection) VAD=0 according to the equation:

$$en_{\log}(i) = \frac{1}{2}\log_2\left(\frac{1}{N}\sum_{n=0}^{N-1} s^2(n)\right), \tag{1.13}$$

where $s(n)$ is the high-pass filtered input speech signal of the current frame i.

[0063] The averaged logarithmic energy is computed by:

$$en_{\log}^{nean}(i) = \frac{1}{8}\sum_{n=0}^{7} en_{\log}(i-n). \tag{1.14}$$

[0064] The averaged logarithmic frame energy is quantized by means of a 6-bit algorithmic quantizer. These 6 bits for the energy index are transmitted in the SID frame.

[0065] In the following, gain control in the parameter domain is described.

[0066] The fixed codebook gain g_c adjusts the level of the synthesized speech in the AMR speech codec, as can be noticed by studying the equation (1.1) and the speech synthesis model shown in Fig. 1.

[0067] The adaptive codebook gain g_p controls the periodicity (pitch) of the synthesized speech, and is limited between [0, 1.2]. As shown in Fig. 1, an adaptive feedback loop transmits the effect of the fixed codebook gain also to the adaptive codebook branch of the synthesis model thereby adjusting also the voiced part of the synthesized speech.

[0068] The speed at which the change in the fixed codebook gain is transmitted to the adaptive codebook branch depends on the pitch delay T and the pitch gain g_p, as illustrated in Fig. 3. The longer the pitch delay and the higher the pitch gain, the longer it takes for the adaptive codebook vector $v(n)$ to stabilize (to reach its corresponding level).

[0069] For real speech signals, the pitch gain and delay vary. However, the simulation with a fixed pitch delay and pitch gain tries to give a rough estimate on the limits to the stabilization time of the adaptive codebook after a change in the fixed codebook gain. The pitch delay is limited in AMR between [18, 143] samples, as in the example too, corresponding to high child and low male pitches, respectively. The pitch gain, however, may have values between [0,1.2]. For zero pitch gain, there is naturally no delay at all. On the other hand, the pitch gain receives values at or above 1 only very short time instants for the adaptive codebook not to go unstable. Therefore, the estimated maximum delay is around few thousand samples, about half a second.

[0070] Fig. 3 shows the response of the adaptive codebook to a step-function (sudden change in g_c) as a function of pitch delay T (integer lag k in Eq. (1.1)) and pitch gain g_p. The output of the scaled fixed codebook, $g_c{*}c(n)$, changes from 0 to 0.3 at time instant 0 samples. The output of the adaptive codebook (and thus also the excitation signal $u(n)$) reaches its corresponding level after 108 to 5430 samples, for the pitch delays T and pitch gains g_p of the example.

[0071] In the highest bit rate mode, 12.2 kbit/s, the fixed codebook gain correction factor γ_{gc} is scalar quantized with 5-bits, giving 32 quantization levels, as shown in Fig. 4. The quantization is nonlinear. The quantization steps are shown in Fig. 5. The quantization step is between 1.2 dB to 2.3 dB.

[0072] The same quantization table is used in the mode 7.95 kb/s. In all other modes, the fixed codebook gain factor is jointly vector quantized with the adaptive codebook gain. These quantization tables are shown in Figs. 6 and 7.

[0073] The lowest mode 4.75 kbit/s uses vector quantization in a unique way. In the mode 4.75 kbit/s the adaptive codebook gains g_p and the correction factors $\hat{\gamma}_{gc}$ are jointly vector quantized every 10 ms with 6 bits, i.e. two codebook gains of two frames and two correction factors are jointly vector quantized.

[0074] Fig. 5 shows a difference between adjacent quantization levels in the quantization table of the fixed codebook gain factor γ_{gc} in the modes 12.2 kbit/s and 7.95 kbit/s. The quantization table is approximately linear between indexes 5 and 28. The quantization step in that range is about 1.2 dB.

[0075] Fig. 6 shows the vector quantization table for the adaptive codebook gain and the fixed codebook gain factor in the modes 10.2, 7.4 and 6.7 kbit/s. The table is printed so that one index value gives both the fixed codebook gain factor and the corresponding (jointly quantized) adaptive codebook gain. As can be seen from Fig. 6, there are approximately 16 levels to choose from for the fixed codebook gain while the adaptive codebook gain remains fairly fixed.

[0076] Fig. 7 shows the vector quantization table for the adaptive codebook gain and the fixed codebook gain factor in the modes 5.90 and 5.15 kbit/s. Again, the table is printed so that one index value gives both the fixed codebook gain factor and the corresponding (jointly quantized) adaptive codebook gain.

[0077] As explained above, the speech level control in the parameter domain must take place by adjusting the fixed codebook gain. To be more specific, the quantized fixed codebook gain correction factor $\hat{\gamma}_{gc}$ is adjusted, which is one of the speech parameters transmitted to the far-end.

[0078] In the following, the relationship between amplification of the fixed codebook gain correction factor and the amplification of the fixed codebook gain is shown. As already shown in Eqs. (1.11) and (1.12), the fixed codebook gain is defined as:

$$\hat{g}_c(n) = \hat{\gamma}_{gc}(n) \cdot 10^{0.05\left[\sum_{i=1}^{4} b_i 20\log_{10}\left(\hat{\gamma}_{gc}(n-i)\right) + \bar{E} - E_I\right]}. \tag{2.1}$$

[0079] If the fixed codebook gain correction factor $\hat{\gamma}_{gc}(n)$ is amplified by β, at subframe n, and is kept unchanged at least for the following four subframes, the new quantized fixed codebook gain becomes:

$$\hat{g}_c^{new}(n) = \beta\hat{\gamma}_{gc}(n) \cdot 10^{0.05\left[\sum_{i=1}^{4} b_i 20\log_{10}\left(\hat{\gamma}_{gc}(n-i)\right) + \bar{E} - E_I\right]} = \beta\hat{g}_c^{old}(n). \tag{2.2}$$

In the next subframe, n+1, the new fixed codebook gain becomes:

$$\hat{g}_c^{new}(n+1) = \beta\hat{\gamma}_{gc}(n+1) \cdot 10^{0.05\left[b_1 20\log_{10}\left(\beta\hat{\gamma}_{gc}((n+1)-1)\right) + \sum_{i=2}^{4} b_i 20\log_{10}\left(\hat{\gamma}_{gc}((n+1)-i)\right) + \bar{E} - E_I\right]} \tag{2.3}$$

$$\hat{g}_c^{new}(n+1) = \beta\hat{\gamma}_{gc}(n+1) \cdot 10^{0.05\left[b_1 20\log_{10}(\beta) + \sum_{i=1}^{4} b_i 20\log_{10}\left(\hat{\gamma}_{gc}((n+1)-i)\right) + \bar{E} - E_I\right]} \tag{2.4}$$

$$\hat{g}_c^{new}(n+1) = \beta\hat{\gamma}_{gc}(n+1) \cdot 10^{0.05[b_1 20\log_{10}(\beta)]} \cdot 10^{0.05\left[\sum_{i=1}^{4} b_i 20\log_{10}\left(\hat{\gamma}_{gc}((n+1)-i)\right) + \bar{E} - E_I\right]} \tag{2.5}$$

$$\hat{g}_c^{new}(n+1) = \beta\hat{\gamma}_{gc}(n+1) \cdot \beta^{b_1} 10^{0.05\left[\sum_{i=1}^{4} b_i 20\log_{10}\left(\hat{\gamma}_{gc}((n+1)-i)\right) + \bar{E} - E_I\right]} \tag{2.6}$$

$$\hat{g}_c^{new}(n+1) = \beta \cdot \beta^{b_1} \hat{g}_c^{old}(n+1)\,. \qquad (2.7)$$

[0080] In the same way, in the following subframes, *n*+2, ..., *n*+4, the amplified fixed codebook gain becomes:

$$\hat{g}_c^{new}(n+2) = \beta \cdot \beta^{b_1} \cdot \beta^{b_2} \hat{g}_c^{old}(n+2) \qquad (2.8)$$

...

$$\hat{g}_c^{new}(n+4) = \beta^{(1+b_1+b_2+b_3+b_4)} \cdot \hat{g}_c^{old}(n+4)\,. \qquad (2.9)$$

[0081] Since the prediction coefficients were given as

$$[b_1\ b_2\ b_3\ b_4] = [0.68\ 0.58\ 0.34\ 0.19],$$

the fixed codebook gain stabilizes after five subframes into a value:

$$\hat{g}_c^{new}(n+4) = \beta^{2.79} \cdot \hat{g}_c^{old}(n+4)\,. \qquad (2.10)$$

[0082] In other words, multiplying the fixed codebook gain factor with β results in multiplication of the fixed codebook gain (and therefore also the synthesized speech) by $\beta^{2.79}$, assuming that β is held constant at least during the next four frames.

[0083] Therefore, e.g. in AMR modes 12.2 kbit/s and 7.95 kbit/s, the minimum change for the fixed codebook gain factor (the minimum quantization step) ±1.2 dB results in ±3.4 dB change in the fixed codebook gain, and hence in the synthesized speech signal, as shown below.

$$\begin{aligned} &20\log_{10}\beta = 1.2\ dB \Leftrightarrow \beta = 1.15 \\ &20\log_{10}\left(\beta^{2.79}\right) = 3.4\ dB \end{aligned} \qquad (2.11)$$

[0084] This ±3.4 dB change in the synthesized speech level takes place gradually, as illustrated in Fig. 8.

[0085] Fig. 8 shows a change in the fixed codebook gain (AMR 12.2 kbit/s), when the fixed codebook gain factor is changed one quantization step (in the linear quantization range) first upwards at subframe 6 and then downwards at subframe 16. The 1.2 dB amplification (or attenuation) of the fixed codebook gain factor amplifies (or attenuates) the fixed codebook gain gradually 3.4 dB during 5 frames (200 samples).

[0086] Consequently, the parameter level gain control of coded speech may be made by changing the index value of the fixed codebook gain factor. That is, the index value in the bit stream is replaced by a new value that gives the desired amplification/attenuation. The gain values corresponding to the index changes for AMR mode 12.2 kbit/s are listed in the table below.

Table I: Parameter level gain values for AMR 12.2 kbit/s.

[00010] Change in the fixed	**[00011]** Resulting amplification/

codebook gain factor index value	attenuation of the speech signal
[00012] ⋮	**[00013]** ⋮
[00014] +4	**[00015]** 13.6 dB
[00016] +3	**[00017]** 10.2 dB
[00018] +2	**[00019]** 6.8 dB
[00020] +1	**[00021]** 3.4 dB
[00022] 0	**[00023]** 0 dB
[00024] -1	**[00025]** - 3.4 dB
[00026] -2	**[00027]** - 6.8 dB
[00028] -3	**[00029]** - 10.2 dB
[00030] -4	**[00031]** - 13.6 dB
[00032] ⋮	**[00033]** ⋮

[00034]

[0087] Next, a search for the correct index for the desired change in the overall gain is described by taking into account the nonlinear nature of the fixed codebook gain factor quantization.

[0088] The new fixed codebook gain factor quantization index corresponding to the desired amplification/attenuation of the speech signal is found by minimizing the error:

$$\left|\beta \cdot \hat{\gamma}_{gc}^{old} - \hat{\gamma}_{gc}^{new}\right|, \qquad (2.12)$$

where $\hat{\gamma}_{gc}^{old}$ and $\hat{\gamma}_{gc}^{new}$ are the old and the new fixed codebook gain correction factors and β is the desired multiplier:

$\beta = \Delta^{j}$, $j = [\ldots -4,-3,..0,..+3,+4,\ldots]$ $\Delta =$ minimum quantization step (1.15 in AMR 12.2 kbit/s)). Note that the speech signal becomes amplified/attenuated with $\beta^{2.79}$.

[0089] Fig. 9 shows the re-quantized levels for cases +3.4, +6.8, +10.2, +13.6 and +17.0 dB signal amplification achieved with the above error minimization procedure. Fig. 10 shows also the quantization levels in cases of signal attenuation. Both figures show the quantization levels for the AMR mode 12.2 kbit/s.

[0090] In Fig. 9 the lowest curve shows the original quantization levels of the fixed codebook gain factor. The second lowest curve shows re-quantized levels of the fixed codebook gain factor in the case of +3.4 dB signal level amplification, and the subsequent curves show re-quantized levels of the fixed codebook gain factor in cases +6.8, +10.2, +13.6 and +17 dB signal level amplification, respectively.

[0091] Fig. 10 shows re-quantized levels of the fixed codebook gain factor in cases: -17, -13.6, ..., -3.4, 0, +3.4, ..., +13.6, +17 dB signal level amplification. The curve in the middle shows the original quantization levels of the fixed codebook gain factor.

[0092] In AMR modes 10.2 kbit/s, 7.40 kbit/s, 6.70 kbit/s, 5.90 kbit/s, 5.15 kbit/s and 4.75 kbit/s, the equation 2.12 is replaced by:

$$\left|\beta \cdot \hat{\gamma}_{gc}^{old} - \hat{\gamma}_{gc}^{new}\right| + weight \cdot \left|g_{p\_new} - g_{p\_old}\right|, \qquad (2.13)$$

where the *weight* is ≥ 1, and g_{p_new} and g_{p_old} are the new and old adaptive codebook gains, respectively.

[0093] In other words, in modes 12.2 kbit/s and 7.95 kbit/s, the new fixed codebook gain factor index is found as the index which minimizes the error given in Eq. (2.12). In modes 10.2 kbit/s, 7.40 kbit/s, 6.70 kbit/s, 5.90 kbit/s, 5.15 kbit/s and 4.75 kbit/s the new joint index of the vector quantized fixed codebook gain factor and adaptive gain is found as the index which minimized the error given in Eq. (2.13). The rationale behind the Eq. (2.13) is to be able to change the fixed codebook gain factor without introducing audible error to the adaptive codebook gain. Fig. 6 shows the vector quantized fixed codebook gain factors and adaptive codebook gains at different index values. From Fig. 6 it can be seen that there is a possibility to change the fixed codebook gain factor without having to change the adaptive codebook gain excessively.

[0094] As mentioned above, in the mode 4.75 kbit/s the adaptive codebook gains g_p and the correction factors $\hat{\gamma}_{gc}$ are jointly vector quantized every 10 ms with 6 bits, i.e. two codebook gains of two subframes and two correction factors are jointly vector quantized. The codebook search is done by minimizing a weighted sum of the error criterion for each of the two subframes. The default values of the weighing factors are 1. If the energy of the second subframe is more than two times the energy of the first subframe, the weight of the first subframe is set to 2. If the energy of the first subframe is more than four times the energy of the second subframe, the weight of the second subframe is set to 2. Despite of these differences, the mode 4.75 kbit/s can be processed with the vector quantization schema described above.

[0095] Thus, according to the above-described embodiment, a new gain index (new index value) minimizing the error between the desired gain $\beta \cdot \hat{\gamma}_{gc}^{old}$ (enhanced first parameter value) and the realized effective gain $\hat{\gamma}_{gc}^{new}$ (new first parameter value) according to Eq. (2.12) or (2.13) is determined according to the quantization tables for the respective modes. The new fixed codebook gain correction factor (and the new adaptive codebook gain in case of modes other than 12.2 kbits/s and 7.95 kbit/s) correspond to the determined new gain index. The old gain index (current index value) representing the old fixed codebook gain correction factor $\hat{\gamma}_{gc}^{old}$ (current first parameter value) (and the old adaptive codebook gain g_{p_old} (current second parameter value) in case of modes other than 12.2 kbits/s and 7.95 kbit/s) then is replaced by the new gain index.

[0096] In the following, alternative methods for providing an improved gain accuracy are described. At first it is illustrated how the total desired gain is formulated in case the gain is not kept constant during five consecutive subframes.

[0097] As described above, in the AMR-codec, the fixed codebook gain is encoded using the fixed codebook gain correction factor γ_{gc}. The gain correction factor is used to scale the predicted fixed codebook gain g'_c to obtain the fixed codebook gain g_c, i.e.

$$g_c = \gamma_{gc} g'_c \Rightarrow \gamma_{gc} = \frac{g_c}{g'_c}.$$

[0100] The fixed codebook gain is predicted as follows:

$$g'_c(n) = 10^{0.05\left[\sum_{i=1}^{4} b_i \, 20 \log_{10}\left(\hat{\gamma}_{gc}(n-i)\right) + \bar{E} - E_I\right]} \qquad \textbf{(3.1)}$$

where $\bar{E}$ is a mode dependent energy value (in dB) and E_I is the fixed codebook excitation energy (in dB).

[0101] To obtain a desired overall signal gain α, the quantified fixed codebook correction factor has to be multiplied by a correction factor gain β. Realized correction factor gains are denoted with $\hat{\beta}(n-i), i > 0$. By amplifying the fixed codebook correction factor $\hat{\gamma}_{gc}(n)$ with $\beta(n)$, at subframe n, the new quantized fixed codebook gain becomes: (Note that the prediction g'_c depends on the history of the correction gains, as shown in Equation 2.14)

$$
\begin{aligned}
&\hat{g}_c^{new}(n) = \beta(n)\hat{\gamma}_{gc}(n)g_c'^{new}(n)\\
&\hat{g}_c^{new}(n) = \beta(n)\hat{\gamma}_{gc}(n)\cdot 10^{0.05\left[\sum_{i=1}^{4} b_i\, 20\log_{10}\left(\hat{\beta}(n-i)\hat{\gamma}_{gc}(n-i)\right)+\overline{E}-E_I\right]}\\
&\hat{g}_c^{new}(n) = \beta(n)\hat{\gamma}_{gc}(n)\cdot 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\hat{\gamma}_{gc}(n-i)\right)+0.05\overline{E}-0.05E_I}\\
&\hat{g}_c^{new}(n) = \beta(n)\hat{\gamma}_{gc}(n)\cdot 10^{\sum_{i=1}^{4} b_i\left(\log_{10}\left(\hat{\beta}(n-i)\right)+\log_{10}\left(\hat{\gamma}_{gc}(n-i)\right)\right)+0.05\overline{E}-0.05E_I}\\
&\hat{g}_c^{new}(n) = \beta(n)\hat{\gamma}_{gc}(n)\cdot 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\gamma}_{gc}(n-i)\right)+0.05\overline{E}-0.05E_I}\\
&\hat{g}_c^{new}(n) = \beta(n)\cdot 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)}\cdot\hat{\gamma}_{gc}(n)\cdot 10^{0.05\left[\sum_{i=1}^{4} b_i\, 20\log_{10}\left(\hat{\gamma}_{gc}(n-i)\right)+\overline{E}-E_I\right]}\\
&\hat{g}_c^{new}(n) = \beta(n)\cdot 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)}\cdot\hat{\gamma}_{gc}(n)g_c'(n)
\end{aligned}
$$

[0102] Therefore, a new prediction, which is obtained using the realized factor gains $\hat{\beta}(n-i)$, can be written as $g_c'^{new} = 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} g_c'$. Furthermore,

$$
\begin{aligned}
&\hat{g}_c^{new}(n) = \hat{\beta}(n)\cdot 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)}\cdot\hat{\gamma}_{gc}(n)g_c'(n)\\
&\hat{g}_c^{new}(n) = 10^{\log_{10}\hat{\beta}(n)}\cdot 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)}\cdot\hat{\gamma}_{gc}(n)g_c'(n)\\
&\hat{g}_c^{new}(n) = 10^{\sum_{i=0}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)}\cdot\hat{\gamma}_{gc}(n)g_c'(n), \qquad b_o = 1\\
&\hat{g}_c^{new}(n) = \alpha g_c(n).
\end{aligned}
$$

i.e., the target correction factor gain for the present subframe can be written as

$$\alpha = 10^{\sum_{i=0}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} \Leftrightarrow \hat{\beta}(n) = \frac{\alpha}{10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)}}.$$

[0103] If $\hat{\beta}(n)$ is kept constant, the overall gain stabilizes after five subframes into a value

$$\alpha = 10^{\sum_{i=0}^{4} b_i \log_{10}\left(\hat{\beta}\right)} = 10^{\log_{10}\left(\hat{\beta}\right)\sum_{i=0}^{4} b_i} = \hat{\beta}^{\sum_{i=0}^{4} b_i} = \hat{\beta}^{2.79} \Leftrightarrow \hat{\beta} = \alpha^{\frac{1}{2.79}} = \text{a},$$

because the prediction coefficients were given as $\mathbf{b} = \left[1, 0.68, 0.58, 0.34, 0.19\right]$.

[0104] Next, a first alternative of the above described gain manipulation is described, which first alternative is referred to as Synthesizing Error Minimization (synthesizing method).

[0105] The algorithm according to the synthesizing method follows as much as possible the original error criteria given for the scalar quantization as

$$E_{SQ} = (g_c - \hat{g}_c)^2 = (g_c - \hat{\gamma}_{gc} g'_c)^2,$$

where E_{SQ} is the fixed codebook quantization error and g_c is the target fixed codebook gain. As mentioned before, the goal is to scale the fixed codebook gain with the desired total gain $g_c^{new} = \alpha\hat{g}_c$. Therefore, for the CDALC (Coded Domain Automatic Level Control) purposes, the target must be scaled by the desired gain, i.e.

$$E_{SQ} = (\alpha\hat{g}_c - \hat{\gamma}_{gc}^{new} g_c'^{new})^2. \qquad (3.2)$$

[0106] In the vector quantization, the pitch gain g_p and the fixed codebook correction factor $\hat{\gamma}_{gc}$ are jointly quantized. In the AMR encoder, the vector quantization index is found by minimizing the quantization error E_{VQ} defined as

$$E_{VQ} = \left\| \mathbf{x} - \hat{g}_p \mathbf{y} - \hat{g}_c \mathbf{z} \right\|,$$

where $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are a target vector, a weighted LP-filtered adaptive codebook vector and a weighted LP-filtered fixed codebook vector, respectively. The error criterion is actually a norm of the perceptually weighted error between the target and the synthesized speech. Following the procedure of the scalar quantization, the target vector is replaced by the scaled version, i.e.

$$E_{VQ} = \left\| (\hat{g}_p \mathbf{y}^{new} + \alpha\hat{g}_c \mathbf{z}) - \hat{g}_p^{new} \mathbf{y}^{new} - \hat{g}_c^{new} \mathbf{z} \right\|. \qquad (3.3)$$

[0107] In the following, the synthesizing method is described for the scalar quantization.

[0108] The derivation of the minimization criterion is started from the Equation 3.2 used in the AMR-encoder and given as:

$$E_{SQ} = \left(\alpha g_c - \hat{\gamma}_{gc}^{new} g_c'^{new}\right)^2.$$

[0109] Unfortunately, there is no direct access to g_c, however it can be approximated by $g_c \approx \hat{\gamma}_{gc} g_c'$ and therefore the first CDALC error criterion for the scalar quantization can be written as

$$\begin{aligned}
E_{SQ} &= \left(\alpha \hat{\gamma}_{gc} g_c' - \hat{\gamma}_{gc}^{new} g_c'^{new}\right)^2 \\
E_{SQ} &= \left(\alpha \hat{\gamma}_{gc} g_c' - \hat{\gamma}_{gc}^{new} 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} g_c'\right)^2 \\
E_{SQ} &= g_c'^{\,2} \left(\alpha \hat{\gamma}_{gc} - 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} \hat{\gamma}_{gc}^{new}\right)^2 \Leftrightarrow \\
E_{SQ'} &= \left|\alpha \hat{\gamma}_{gc} - 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} \hat{\gamma}_{gc}^{new}\right|
\end{aligned} \qquad (3.4)$$

where $\hat{\beta}(n-i)$ is the realized correction factor gain for the subframe $(n-i)$, i.e.

$$\hat{\beta}(n-i) = \frac{\hat{\gamma}_{gc}^{new}(n-i)}{\hat{\gamma}_{gc}(n-i)}.$$

[0110] This error criterion is simple to evaluate and only the fixed codebook correction factor has to be decoded. Furthermore, four previous realized correction factor gains have to be kept in the memory.

[0111] Next, the synthesizing method is described for the vector quantization.

[0112] For the vector quantization case the error criterion used in the AMR-encoder is more complicated, since the synthesis filters are used. In view of the fact that there is no direct access to the target $\mathbf{x}$, it is approximated by $\hat{g}_p\mathbf{y} + \hat{g}_c\mathbf{z}$. Thus, the error minimization with CDALC becomes:

$$\begin{aligned}
E_{VQ} &= \left\| \mathbf{x}^{new} - \hat{g}_p^{new} \mathbf{y}^{new} - \hat{g}_c^{new} \mathbf{z} \right\| \\
E_{VQ} &= \left\| (\hat{g}_p \alpha \mathbf{y} + \alpha \hat{g}_c \mathbf{z}) - \hat{g}_p^{new} \alpha \mathbf{y} - \hat{g}_c^{new} \mathbf{z} \right\| \\
E_{VQ} &= \left\| \left(\hat{g}_p - \hat{g}_p^{new} \right) \alpha \mathbf{y} + \left(\alpha \hat{g}_c - \hat{g}_c^{new} \right) \mathbf{z} \right\| \\
E_{VQ} &= \left\| \left(\hat{g}_p - \hat{g}_p^{new} \right) \alpha \mathbf{y} + \left(\alpha \ \hat{\gamma}_{gc} g_c' - \hat{\gamma}_{gc}^{new} g_c'^{new} \right) \mathbf{z} \right\| \\
E_{VQ} &= \left\| \left(\hat{g}_p - \hat{g}_p^{new} \right) \alpha \mathbf{y} + g_c' \left(\alpha \hat{\gamma}_{gc} - \hat{\gamma}_{gc}^{new} 10^{\sum_{i=1}^{4} b_i \log_{10}\left(\hat{\beta}(n-i)\right)} \right) \mathbf{z} \right\|.
\end{aligned} \qquad (3.5)$$

[0113] In addition to decoding the gains, both codebook vectors have to be decoded and filtered with the LP-synthesis filter. Therefore, LP-synthesis filter parameters have to be decoded. This means that basically all the parameters have to be decoded. In the AMR-encoder the codebook vectors are also weighted by a specific weighting filter, but this was not done for this CDALC error criterion.

[0114] Next, a second alternative of the gain manipulation is described, which second alternative is referred to as Quantization Error Minimization with Memory (memory method).

[0115] This criterion minimizes quantization error while taking in account the history of the previous correction factors. In case of scalar quantization the error criterion is the same as in the first alternative, i.e. the error function to be minimized will be the same as in Equation 3.4. But for the vector quantization the error function becomes little easier to evaluate.

Vector Quantization

[0116] Starting from the error function derived for the first alternative and given in Equation 3.5, minimizing the error of the sum of two components will require decoding the $\mathbf{y}$ and $\mathbf{z}$ vectors. Practically this means that the whole signal has to be decoded. Instead of minimizing the norm of the error vector, the error can be approximated by the sum of two error components (which would be the case if both vectors y and z are parallel to each other), namely the pitch gain error and the fixed codebook gain error. Combining these components using the Euclidean norm, the new error criteria can be written as:

$$E_{VQ'}=\sqrt{\left\|\left(\hat{g}_p-\hat{g}_p^{new}\right)\alpha\mathbf{y}\right\|^2+\left\|g_c'\left(\alpha\hat{\gamma}_{gc}-\hat{\gamma}_{gc}^{new}10^{\sum_{i=1}^{4}b_i\log_{10}\left(\hat{\beta}(n-i)\right)}\right)\mathbf{z}\right\|^2}$$

$$E_{VQ'}=\sqrt{\left|\hat{g}_p-\hat{g}_p^{new}\right|^2\left\|\alpha\mathbf{y}\right\|^2+\left|\alpha\hat{\gamma}_{gc}-\hat{\gamma}_{gc}^{new}10^{\sum_{i=1}^{4}b_i\log_{10}\left(\hat{\beta}(n-i)\right)}\right|^2 g_c'^2\left\|\mathbf{z}\right\|^2}\Rightarrow \qquad (3.6)$$

$$E_{VQ''}=\left|\hat{g}_p-\hat{g}_p^{new}\right|^2\left(\frac{\alpha\left\|\mathbf{y}\right\|}{g_c'\left\|\mathbf{z}\right\|}\right)^2+\left|\alpha\hat{\gamma}_{gc}-\hat{\gamma}_{gc}^{new}10^{\sum_{i=1}^{4}b_i\log_{10}\left(\hat{\beta}(n-i)\right)}\right|^2.$$

[0117] The sum of the previous equation (Equation 3.5) is divided into two components. However, the synthesized codebook vectors still exist in the pitch gain error scaling term $\left(\frac{\alpha\left\|\mathbf{y}\right\|}{g_c'\left\|\mathbf{z}\right\|}\right)^2$. Due to the synthesis, the pitch gain error scaling term is complicate to compute. If it is computed, it would be more efficient to use the synthesization error minimization criterion described in the first alternative. To get rid of the synthesis-procedure, the term $\frac{\left\|\mathbf{y}\right\|}{\left\|\mathbf{z}\right\|}$ is replaced by the constant pitch gain error weight w_{g_p}. The pitch gain error weight has to be chosen carefully. If the weight is chosen to be too big, the signal level will not change at all, since the lowest error is found by choosing $g_p^{new}=g_p$. On the other hand, a small weight will guarantee the desired codebook gain α, but it will give no guarantees for g_p, i.e.

$$w_{g_p}\rightarrow 0\Rightarrow \text{minimization of term}\left|\alpha\hat{\gamma}_{gc}-\hat{\gamma}_{gc}^{new}10^{\sum_{i=1}^{4}b_i\log_{10}\left(\hat{\beta}(n-i)\right)}\right|^2.$$

$$w_{g_p}\rightarrow\infty\Rightarrow \text{minimization of term}\left|g_p^{old}-g_p^{new}\right|^2$$

[0118] This algorithm using fixed pitch gain weight requires decoding (finding a value according to the received quantization index) of both the pitch gain and the correction factor ($\hat{\gamma}_{gc}$) and also reconstructing of the fixed codebook gain prediction g_c'. To be able to construct the prediction, the fixed codebook vector has to be decoded. Furthermore, the integer pitch lag is needed for the pitch sharpening of the fixed codebook excitation. The energy of the fixed codebook excitation is required for the prediction (see Equation 3.1). If necessary, the prediction can be included in the fixed weight, i.e. $w_{g_p} = \frac{\|\mathbf{y}\|}{g_c'\|\mathbf{z}\|}$. After that there is no need to decode the fixed codebook vector. Presumably, it would not affect much in performance. On the other hand, the energy of the fixed codebook excitation can be estimated, since it is fairly fixed. This allows the creation of a prediction without decoding the fixed codebook vector.

[0119] The range of the terms $\frac{\|\mathbf{y}\|}{\|\mathbf{z}\|}$ and $\frac{\|\mathbf{y}\|}{g_c'\|\mathbf{z}\|}$ are demonstrated in Figs. 11 and 12 with male and child speech samples using AMR mode 12.2 kbit/s. The value depends strongly on the energy of the signal. Hence, it would be beneficial to make the pitch gain error weight w_{g_p} adaptive instead of using a constant value. For example, the value may be determined using short time signal energy.

[0120] Fig. 13 shows a flow chart generally illustrating the method of enhancing a coded audio signal comprising coded speech and/or coded noise according to the invention. The coded audio signal comprises indices which represent speech parameters and/or noise parameters which comprise at least a first parameter for adjusting a first characteristic of the audio signal, such as the level of synthesized speech and/or noise.

[0121] In step S1 in Fig. 13 a current first parameter value is determined from an index corresponding to at least the first parameter, e.g. the fixed codebook gain correction factor $\hat{\gamma}_{gc}$. In step S2 the current first parameter value is adjusted, e.g. multiplied by a, in order to achieve an enhanced first characteristic, thereby obtaining an enhanced first parameter value a$\cdot\hat{\gamma}_{gc}^{old}$. Finally, in step S3 a new index value is determined from a table relating index values to at least first parameter values, e.g. a quantization table, such that a new first parameter value corresponding to the new index value substantially matches the enhanced first parameter value.

[0122] According to the above-described embodiment, a new index value for a$\cdot\hat{\gamma}_{gc}^{old}$ is searched such that the equation $\left|a\cdot\hat{\gamma}_{gc}^{old}-\hat{\gamma}_{gc}^{new}\right|$ is minimized, $\hat{\gamma}_{gc}^{new}$ being the new first parameter value corresponding to the searched new index value.

[0123] Moreover, according to the present invention, a current second parameter value may be determined from the index further corresponding to a second parameter such as the adaptive codebook gain controlling a second characteristic of speech. In this case, the new index value is determined from the table further relating the index values to second parameter values, e.g. a vector quantization table, such that a new second parameter value corresponding to the new index value substantially matches the current second parameter value.

[0124] According to the above-described embodiment, a new index value for a$\cdot\hat{\gamma}_{gc}^{old}$ and g_{p_old} is searched such that the equation $\left|a\cdot\hat{\gamma}_{gc}^{old}-\hat{\gamma}_{gc}^{new}\right|+weight\cdot\left|g_{p\_new}-g_{p\_old}\right|$ is minimized. g_{p_new} is the new second parameter value corresponding to the new index value.

[0125] "*weight*" can be ≥ 1, so that the new index value is determined from the table such that substantially matching the current second parameter value has precedence.

[0126] Fig. 14 shows a schematic block diagram illustrating an apparatus 100 for enhancing a coded audio signal according to the invention. The apparatus receives a coded audio signal which comprises indices which represent speech and/or noise parameters which comprise at least a first parameter for adjusting a first characteristic of the audio signal. The apparatus comprises a parameter value determination block 11 for determining a current first parameter value from an index corresponding to at least the first parameter, an adjusting block 12 for adjusting the current first parameter value in order to achieve an enhanced first characteristic, thereby obtaining an enhanced first parameter value, and an index value determination block 13 for determining a new index value from a table relating index values to at least first parameter values, such that a new first parameter value corresponding to the new index value substantially matches the enhanced first parameter value.

[0127] The parameter value determination block 11 may further determine a current second parameter value from the index further corresponding to a second parameter, and the index value determination block 13 may then determine the new index value from the table further relating the index values to second parameter values, such that a new second parameter value corresponding to the new index value substantially matches the current second parameter value. Thus, the index value is optimized simultaneously for both the first and second parameters.

[0128] The index value determination block 13 may determine the new index value from the table such that substantially matching the current second parameter value has precedence.

[0129] The apparatus 100 may further include replacing means for replacing a current value of the index corresponding to the at least first parameter by the determined new index value, and output enhanced coded speech containing the new index value.

[0130] Referring to Figs. 13 and 14, the first parameter value may be the background noise level parameter value which is determined and adjusted and for which a new index value is determined in order to adjust the background noise level.

[0131] Alternatively, the second parameter value may be the background noise level parameter the index value of which is determined in accordance with the adjusted speech level.

[0132] As discussed beforehand, the speech level manipulation requires also manipulating the background noise level parameter during speech pauses in DTX.

[0133] According to the AMR codec, the background noise level parameter, the averaged logarithmic frame energy, is quantized with 6 bits. The comfort noise level can be adjusted by changing the energy index value. The level can be adjusted in 1.5 dB, so finding a suitable comfort noise level corresponding to the change of the speech level is possible.

[0134] The evaluated comfort noise parameters (the average LSF (Line Spectral Frequency) parameter vector $\mathbf{f}^{mean}$ and the averaged logarithmic frame energy $en_{\log}^{mean}$) are encoded into a special frame, called a Silence Descriptor (SID) frame for transmission to the receiver side. The parameters give information on the level ($en_{\log}^{mean}$) and the spectrum ($\mathbf{f}^{mean}$) of the background noise. More details can be found in 3GPP TS 26.093 V4.0.0 (2001-03), "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory Speech Codec speech processing functions; AMR speech codec; Source controlled rate operation (Release 6)".

[0135] The frame energy is computed for each frame marked with Voice Activity Detector VAD=0 according to the equation:

$$en_{\log}(i) = \frac{1}{2}\log_2\left(\frac{1}{N}\sum_{n=0}^{N-1} x^2(n)\right),$$

where x is the HP-filtered input speech signal of the current frame i. The averaged logarithmic energy, which will be transmitted, is computed by:

$$en_{\log}^{mean}(i) = \frac{1}{8}\sum_{m=0}^{7} en_{\log}(i-m).$$

[0136] The averaged logarithmic energy is quantized by means of a 6 bit algorithmic quantizer. Quantization is performed using quantization function, as defined in 3GPP TS 26.104 V4.1.0 2001-06, "AMR Floating-point Speech Codec C-source".

$$index = \left\lfloor \left(en_{\log}^{mean}(i) + 2.5\right)\cdot 4 + 0.5 \right\rfloor,$$

where the value of the index is restricted to a range $[0...63]$, i.e. in a range of 6 bits.

[0137] The index can be computed using base 10 logarithm as follows:

$$index = \left\lfloor \left(en_{\log}^{mean}(i) + 2.5\right) \cdot 4 + 0.5 \right\rfloor = \left\lfloor 4 \cdot en_{\log}^{mean}(i) + 10.5 \right\rfloor$$

$$index = \left\lfloor 4\frac{1}{2}\frac{\log_{10} en^{mean}(i)}{\log_{10} 2} + 10.5 \right\rfloor = \left\lfloor 2\frac{1}{10}\frac{10\log_{10} en^{mean}(i)}{\log_{10} 2} + 10.5 \right\rfloor,$$

$$index \approx \left\lfloor \frac{1}{1.5} 10\log_{10} en^{mean}(i) + 10.5 \right\rfloor$$

where $10\log_{10} en^{mean}(i)$ is the energy in decibels. Therefore, it is shown that one quantization step corresponds to approximately 1.5 dB.

[0138] In the following the gain adjustment of the comfort noise parameters is described.

[0139] Since an energy parameter is transmitted, the signal energy can be manipulated directly by modifying the energy parameters. As shown above, one quantization step equals to 1.5 dB. Assuming that all eight frames of a SID update interval will be scaled by α, the new index can be found as follows

$$index^{new} = \left\lfloor \left(en_{\log}^{mean}(i) + \frac{1}{2}\log_2 \alpha^2 + 2.5\right) \cdot 4 + 0.5 \right\rfloor = \left\lfloor 4 \cdot en_{\log}^{mean}(i) + 10.5 + 4\log_2 \alpha \right\rfloor.$$

Because the old index was as

$$index = \left\lfloor 4 \cdot en_{\log}^{mean}(i) + 10.5 \right\rfloor,$$

the new index can be approximated by

$$index^{new} \approx \left\lfloor 4\log_2 \alpha \right\rfloor + index.$$

[0140] Referring back to Figs. 13 and 14, a parameter value to be adjusted may be the comfort noise parameter value. Accordingly, a new index value $index^{new}$ is determined as mentioned above. In other words, a current background noise parameter index value *index* may be detected, and a new background noise parameter index value $index^{new}$ may be determined by adding $\lfloor 4\log_2 \alpha \rfloor$ to the current background noise parameter index value *index*, wherein α corresponds to the enhancement of the first characteristic represented by the first speech parameter.

[0141] The level of the synthesized speech signal can be adjusted by manipulating the fixed codebook gain factor index, as shown previously. While being a measure of prediction error, the fixed codebook gain factor index does not discover the level of the speech signal. Therefore, to control the gain manipulation, i.e. to determine whether the level should be changed, the speech signal level must be first estimated.

[0142] In TFO, the six or seven MSB of the PCM speech samples (not compressed) are transmitted to the far end unchanged, to facilitate a seamless TFO interruption. These six or seven MSB can be used to estimate the speech level.

[0143] If these PCM speech samples are unavailable, the coded speech signal must be at least partially decoded (post-filtering is not necessary) to estimate the speech level.

[0144] Alternatively, there is the possibility of using a fixed gain, thereby avoiding a complete decoding. Fig. 15 shows a block diagram illustrating a scheme with the possibility of using a constant gain in the gain manipulation described above. In this case, decoding PCM signals out of the codec signal for using the PCM signals in the gain estimation (i.e. speech level estimation) is not required. The speech may be coded with e.g. AMR, AMR-WB (AMR WideBand), GSM FR, GSM EFR, GSM HR speech codecs.

[0145] Fig. 16 shows a high level implementation example of the present invention in an MGW (Media GateWay) of the 3G network architecture. For example, the present invention may be implemented in a DSP (Digital Signal Processor) of the MGW. However, it is to be noted that the implementation of the invention is not limited to an MGW.

[0146] As shown in Fig. 16, coded speech is fed to the MGW. The coded speech comprises at least one index corresponding to a value of a speech parameter which adjusts the level of synthesized speech. This index may also indicate a value of another speech parameter which is affected by the speech parameter for adjusting the level of synthesized speech. For example, this other speech parameter adjusts the periodicity or pitch of the synthesized speech.

[0147] In a VED (Voice Enhancement Device) shown in Fig. 16, the index is controlled so as to adjust the level of the speech to a desired level. A new index indicating values of the speech parameters affecting the level of the speech, such as the fixed codebook gain factor and adaptive codebook gain, is determined by minimizing an error between the desired level and the realized effective level. As a result, the new index is found which indicates values of the speech parameters realizing the desired level of speech. The original index is replaced by the new index and enhanced coded speech is output.

[0148] It is to be noted that the partial decoding of speech shown in Fig. 16 relates to controlling means for determining a current level of speech to decide whether the level should be adjusted.

[0149] The above described embodiments of the present invention may not only be utilized in level control itself, but also in noise suppression and echo control (nonlinear processing) in the coded domain. Noise suppression can utilize the above technique by e.g. adjusting the comfort noise level during speech pauses. Echo control may utilize the above technique e.g. by attenuating the speech signal during echo bursts.

[0150] The present invention is not intended to be limited only to TFO and TrFO voice communication and to voice communication over packet-switched networks, but rather to comprise enhancing coded audio signals in general. The invention finds application also in enhancing coded audio signals related e.g. to audio/speech/multimedia streaming applications and to MMS (Multimedia Messaging Service) applications.

[0151] It is to be understood that the above description is illustrative of the invention and is not to be construed as limiting the invention. Various modifications and applications may occur to those skilled in the art without departing from the scope of the invention as defined by the appended claims.